



# Research Reproducibility in Political Science

**Dr Nicole Janz**

School of Politics and IR



Dr Nicole Janz  
Assist. Prof. in International Relations



The University of  
**Nottingham**

UNITED KINGDOM • CHINA • MALAYSIA

## Replication



## Corruption



## Human Rights



Political Science Replication Blog



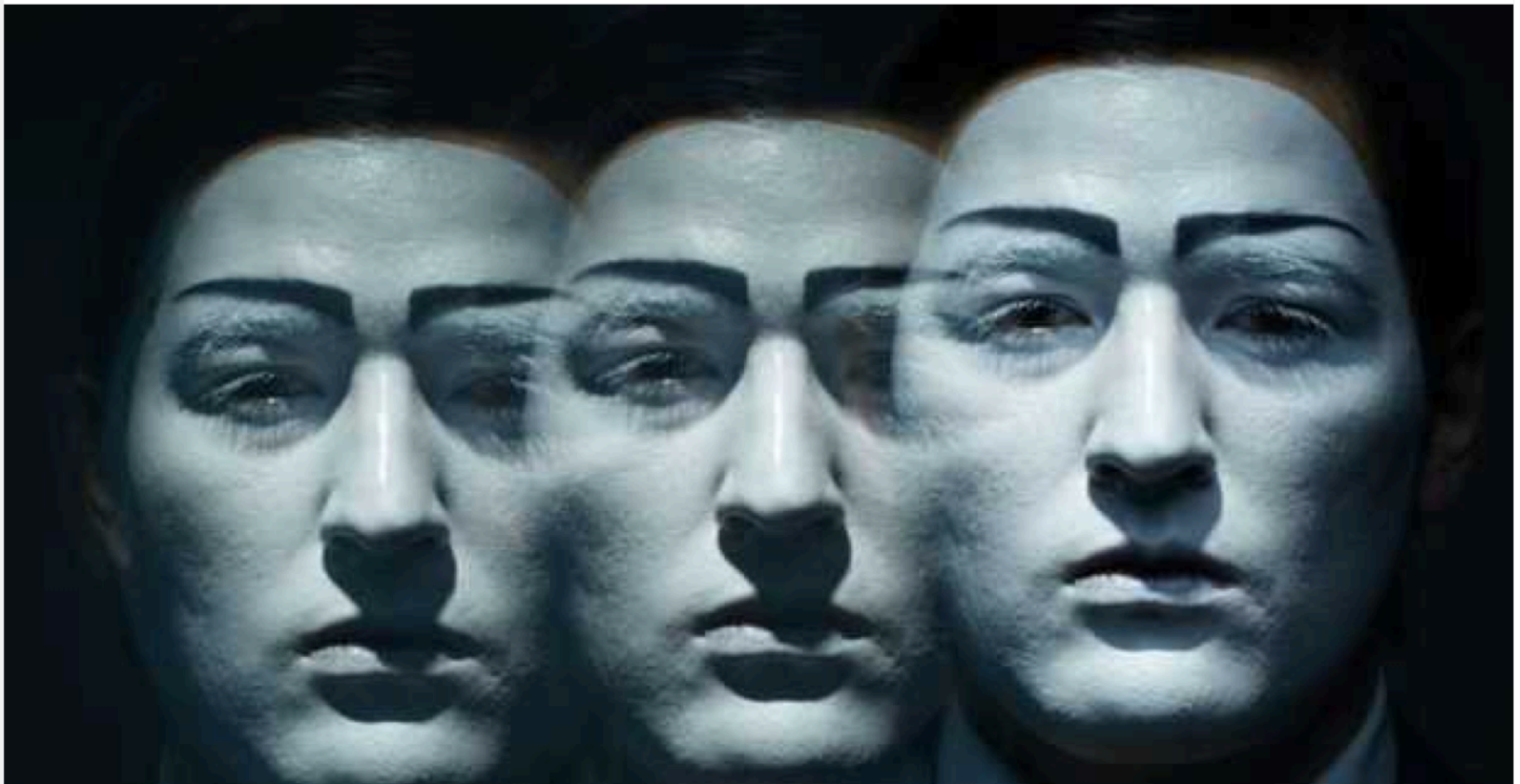
@polscireplicate

# More than half of psychology papers are not reproducible

Initiative to replicate findings of 100 prominent studies raises further questions about health of discipline

August 27, 2015

By [Paul Jump](#) Twitter: [@PaulJump](#)



# The Case of the Amazing Gay-Marriage Data: How a Graduate Student Reluctantly Uncovered a Huge Scientific Fraud

By Jesse Singal



NEW YORK



How can we  
establish  
trust in science?

# Data Access & Research Transparency (DA-RT)

**Data  
transparency**

**Providing full  
access to data  
itself**

**Analytic  
transparency**

**Information  
about data  
analysis**

**Production  
transparency**

**Process of data  
collection**

**Data  
transparency**

**Providing full  
access to data  
itself**

**Analytic  
transparency**

**Information  
about data  
analysis**

**Production  
transparency**

**Process of data  
collection**

## **Quantitative research**

Upload datasets  
used for analysis

Code for models  
(SPSS, STATA, R)

Good methods  
section in paper

Provide or describe  
raw data & variable  
codings

# University training







# Replication exercises

# Bringing the Gold Standard into the Classroom: Replication in University Teaching<sup>1</sup>

NICOLE JANZ

*University of Cambridge*

Reproducibility is held to be the gold standard for scientific research. The credibility of published work depends on being able to replicate the results. However, there are few incentives to conduct replication studies in political science. Replications are difficult to conduct, time-consuming, and hard to publish because of a presumed lack of originality. This article sees a solution in a profound change in graduate teaching. Universities should introduce replications as class assignments in methods training or invest in new stand-alone replication workshops to establish a culture of replication and reproducibility. This article will

# Why should you replicate?

Learn Statistics

- Real life data
- Author decisions
- Bugs included
- More fun than textbook

Reproducibility  
routine

- When are published results really reproducible?

Publish

- Add value
- Publish faster

# Use terminology accepted in **your** field

Political Science (see King 2003)

Duplication	Replication
Verify research results  <i>exact same data set</i> <i>exact same methods</i>	Test the robustness of the original research results  <i>new data</i> <i>new models</i>



# Practical steps in a replication study

- 1 Select paper
  - 2 Access data & code
  - 3 Identify each variable
  - 4 Reproduce tables, figures
  - 5 Compare
- 
- 2-3 weeks
- 3-4 weeks

If you got to this point, you completed a **duplication**.

# Practical steps in a replication study (II)

## 6 Add value

- new data
- new variables
- new model specifications
- theoretical contributions



4-6 weeks

## 7 Compare

## 8 Get feedback from peers

## 9 Journal submission



months

You now completed a full **replication!**

# Comparing your results with the original study

Clarify with **precision** the extent to which you were able to replicate the author's results.

Gary King (2006)

- Exact same data and methods: results cannot be **uplicated**
- **New** data, experiments, models, methods: describe **exactly** at which **step** the results changed and why
- **Different measures** of a concept can naturally yield different results
- **Different lab conditions** may lead to different results

# Communicating failed replications



**Be professional!**



# What replicators write



“We ... find that coding errors, selective exclusion of available data, and unconventional weighting of summary statistics lead to **serious errors**”  
(Herndon et al. 2013)

“If we cannot even reproduce the original results using the same publicly available data, there is **no need for further commentary.**” (Miller et al, 2001)

# How original authors often **respond**

“less realistic”, “inconsistent with the substantive literature,” and “**of limited utility**” (Mansfield, Milner, and Rosendorff 2002)



“fundamentally **flawed**”  
(Peffley, Knigge, and Hurwitz 2001)

“statistical, computational, and reporting errors that **invalidate its conclusions**” (Gerber and Green 2005:301).

# Voting Costs and Voter Turnout in Competitive Elections

Bernard Fraga<sup>1</sup> and Eitan Hersh<sup>2,\*</sup>

Our estimation approach builds off of the methodology and data used by Gomez et al. (2007) ..., adding measures of electoral closeness in order to focus on how the randomly assigned cost (rain) has a different impact depending on the electoral environment.

same way even to rain, then serious doubt should meet claims that voters will react

# Questioning the Effect of Nuclear Weapons on Conflict

Journal of Conflict Resolution

00(0) 1-19

© The Author(s) 2013

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0022002713499718

jcr.sagepub.com



**Mark S. Bell<sup>1</sup> and Nicholas L. Miller<sup>1</sup>**

... we analyze a dyad-year data set (used by Rauchhaus 2009) to examine whether existing findings on the effect of symmetric nuclear weapons possession on conflict are robust to the improvements noted above. We find that once pre-nuclear dyadic conflict is controlled for, symmetric nuclear dyads are not more likely to experience low-level conflict.





## Can inflation expectations be measured using commodity futures prices?

Rasheed Saleuddin<sup>a,c,\*</sup>, D'Maris Coffman<sup>b</sup>

<sup>a</sup> West Face Capital, Canada  
<sup>b</sup> The Bartlett, University College London, London, UK  
<sup>c</sup> University College London, London, UK

### ARTICLE INFO

**Article history:**  
 Received 3 June 2016  
 Received in revised form 4 December 2017  
 Accepted 5 January 2018  
 Available online 3 February 2018

**JEL:**  
 E3  
 Keywords:  
 Commodity futures  
 Inflation expectations  
 Normal backwardation

### ABSTRACT

This paper reexamines the use of US commodity futures price data to show that the US deflation of 1929–1932 was at best no more than partially anticipated by economic actors. By focusing on the expected real interest rate, previous studies provide some empirical support for explanations of the Great Depression that are not exclusively monetary in nature. However, these studies did not consider the context and the market microstructures from which the data was sourced. Our analysis suggests that it is more likely that agricultural commodity markets adjusted to deflationary expectations by the end of 1930. Evidence from commodities futures markets, such as the Chicago Board of Trade, therefore should not be used to critique the Keynesian challenge to the classical monetarist explanation of the Great Depression.

© 2018 Elsevier B.V. All rights reserved.

### 1. Introduction

In economics, there remains a significant controversy about the causes of the length and depth of the Great Depression. If any, of the severe deflation from 1930 to 1933, Friedman and Schwartz (1963) argue for a structural cause. Other authors, while agreeing partially with their point to other important transmission effects, such as banks as credit information holders (e.g., Bernanke, 1933). Finally, there are those, mostly Keynesians, a monetarist explanation, and focus on the role of securitization (Termin, 1976). There are significant problems with a monetarist explanation. Specifically, Friedman and Schwartz cannot fully explain the decline in the velocity of the period, nor are some convinced that the direct runs from money supply declines to real output (e.g., 1976). An additional problem with the Friedman–Schwartz is that interest rates during the Great Depression were low to indicate the severe monetary contraction. The Federal Reserve is accused of accelerating (Termin, 1976) argument, voiced by Brunner (1981) is that, as expected deflation after 1929, real interest rates ex

\* Corresponding author.  
 E-mail address: rsaleuddin@ucl.ac.uk (R. Saleuddin).

https://doi.org/10.1016/j.sced.2018.01.002  
 0950-0804/© 2018 Elsevier B.V. All rights reserved.



Psychophysiology, 52 (2015), 359–366. Wiley Periodicals, Inc. Printed in the USA.  
 Copyright © 2014 The Authors. Psychophysiology published by Wiley Periodicals, Inc. on behalf of Society for Psychophysiological Research  
 DOI: 10.1111/psyp.12336

## Near-wins and near-losses in gambling: A behavioral and facial EMG study

YIN WU,<sup>a</sup> ERIC VAN DUKE,<sup>b</sup> AND LUKE CLARK<sup>a,c</sup>

<sup>a</sup>Behavioral and Clinical Neuroscience Institute, Department of Psychology, University of Cambridge, Cambridge, UK  
<sup>b</sup>Department of Social and Organizational Psychology and Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands  
<sup>c</sup>Centre for Gambling Research at UBC, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada

### Abstract

This study investigated responses to near-wins (i.e., nonwin outcomes that were close to a major win, and their counterpart, near-losses (nonwin outcomes that were proximal to a major loss) in a decision-making task, measuring (a) lack ratings, (b) adjustment of bet amount, and (c) facial muscle reactivity at zygomatics and corrugator sites. Compared to full-misses, near-wins decreased self-perceived lack and near-losses increased self-perceived lack, consistent with the effects of upward versus downward counterfactual thinking, respectively. Wins and losses both increased zygomatic reactivity, and losses selectively enhanced corrugator reactivity. Near-wins heightened zygomatics activity, but did not affect corrugator activity, thus showing a similar response pattern to actual wins. There were no significant facial EMG effects of near-losses. We infer that near-wins engender some appetitive processing, despite their objective nonwin status.

**Descriptors:** Electromyography, Risk taking, Cognitive distortion, Near-miss, Gambling

Gambling is a widespread form of entertainment where a monetary wager is placed upon the uncertain prospect of a larger monetary win. Its allure can provide insight into the psychological mechanisms of human decision making. Previous research has shown that near-wins—nonwin outcomes that are proximal to a jackpot—foster persistent play (Côté, Caron, Aubert, Desrochers, & Ladouceur, 2003; Kassinove & Schare, 2001) and increase motivational ratings (Clark, Lawrence, Astley-Jones, & Gray, 2009). Slot machine near-wins were perceived as being “closer” to wins than to losses (Dymond et al., 2014). Using functional magnetic resonance imaging, near-wins were also found to increase neural signal in brain reward circuitry that overlapped with the jackpot wins (Chase & Clark, 2010; Clark et al., 2009). Nevertheless, near-wins also have a negative emotional component; for example, they are rated as significantly less pleasant than full-miss outcomes (Clark, 2010; Clark et al., 2009, 2013; Qi, Ding, Song, & Yang, 2011).

This work was completed within the University of Cambridge Behavioral and Clinical Neuroscience Institute (director: TW Robbins), supported by a consortium award from the Medical Research Council (MRC Ref G100183) and Wellcome Trust (WT Ref 093875/Z/02). YW was supported by a Chinese Scholarship Council–Cambridge International Scholarship and the Terence Stephenson in Biological Sciences from Downing College, Cambridge. The Centre for Gambling Research at UBC is supported by an award from the British Columbia Lottery Corporation and the British Columbia Government. We thank Dr. Aleksandra Kogan for his advice on statistics.

Address correspondence to: Yin Wu, Department of Psychology, University of Cambridge, Cambridge, CB2 3EQ, UK. E-mail: yw330@cam.ac.uk  
 This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

[The copyright line for this article was changed on 29 September 2014 after original online publication.]

Psychophysiology may provide a useful tool for further characterizing the bivalent emotional response to these events. Past work has shown that near-wins increase electrodermal activity (EDA) and heart rate acceleration, in comparison to full-miss outcomes (Clark, Crooks, Clarke, Aitken, & Dunn, 2012; Clark et al., 2013; Dixon et al., 2011). EDA is a sensitive marker of physiological arousal that responds to both aversive and appetitive stimuli, and thus offers limited valence specificity (Dewson, Schell, & Fillion, 2007; Lobbestael, Arntz, & Wiers, 2008). Phasic heart rate changes do show some differences by valence (e.g., Bradley, Codrington, Cuthbert, & Lang, 2001), but this is a complex multiphase response that also varies across individuals (Hodes, Cook, & Lang, 1985). Facial muscle activity offers an alternative probe of stimulus-evoked emotionality with superior valence differentiation, with zygomaticus activity (recorded on the cheek) linked to appetitive processing, and corrugator supercilii activity (recorded on the eyebrow) scaling with aversive processing (Cacioppo, Petty, Losch, & Kim, 1986; Lang, Greenwald, Bradley, & Hamm, 1993). The present study employed facial electromyography (EMG) at these two sites to better decompose the bivalent emotional nature of gambling “near” events.

Little attention has been paid to the natural counterpart to the near-win, the “near-loss.” In a decision with the possibility of losing money, what happens if you discover that you narrowly missed a major loss? While these events are less ubiquitous in gambling behavior, they do occur across many areas of day-to-day decision making, for example, when we narrowly miss an accident, or traffic jam, and they have received some attention in occupational psychology. For example, narrowly avoiding a great disaster lowered the future perceived risk of that event occurring, and increased risk choice (Dillon & Tinsley, 2008). The present study

## Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness

Alexander Coppock<sup>a</sup>

### Abstract

A field experiment carried out by Butler and Nickerson (Butler, D. M., and Nickerson, D. W. (2011). Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quarterly Journal of Political Science* 6, 55–83) shows that New Mexico legislators changed their voting decisions upon receiving reports of their constituents' preferences. The analysis of the experiment did not account for the possibility that legislators may share information, potentially resulting in spillage effects. Working within the analytic framework,

proposed by Bower and indirect treatment to be twice as large

**Keywords:** Field exper

### INTRODUCTION

Butler and Nickerson testing the response of opinion and legislative similarity of preference other possible explanation the effect of learning providing some legislative The headline finding behavior upon acqu

The estimates of on an assumption o

The author is grateful to Dolan, Albert Fang, and Butler and David Nickerson<sup>a</sup> Columbia University, N

© The Experimental Re

### Article

Journal of Conflict Resolution  
 00(0) 1–19

© The Author(s) 2013

Reprints and permission:  
 sagepub.com/journalsPermissions.nav  
 DOI: 10.1177/0022002713499718  
 jcr.sagepub.com



## Questioning the Effect of Nuclear Weapons on Conflict

Mark S. Bell<sup>1</sup> and Nicholas L. Miller<sup>1</sup>

### Abstract

We examine the effect of nuclear weapons on interstate conflict. Using more appropriate methodologies than have previously been used, we find that dyads in which both states possess nuclear weapons are not significantly less likely to fight wars, nor are they significantly more or less belligerent at low levels of conflict. This stands in contrast to previous work, which suggests nuclear dyads are some 2.7 million times less likely to fight wars. We additionally find that dyads in which one state possesses nuclear weapons are more prone to low-level conflict (but not more prone to war). This appears to be because nuclear-armed states expand their interests after nuclear acquisition rather than because nuclear weapons provide a shield behind which states can aggress against more powerful conventional-armed states. This calls into question conventional wisdom on the impact of nuclear weapons and has policy implications for the impact of nuclear proliferation.

### Keywords

nuclear weapons, nuclear proliferation, international conflict

What effect do nuclear weapons have on interstate conflict behavior? Do nuclear weapons bolster the defense, deterring aggression and making states more secure? Or do nuclear weapons embolden the states that possess them, leading to conflicts that

<sup>1</sup> Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, USA

### Corresponding Author:

Mark S. Bell, Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.  
 Email: markbell@mit.edu



# Transparent workflows

# 4 Steps to Transparency

**Fixed Folder  
Structure**

**Comment your code**

**Clear methods  
section in paper**

**Share materials**

## **Fixed folder structure**

Decide on a template structure for each project

Never alter raw data

# Project TIER

## Teaching Integrity in Empirical Research

### **Main**

Readme.txt  
Paper

**/Data**

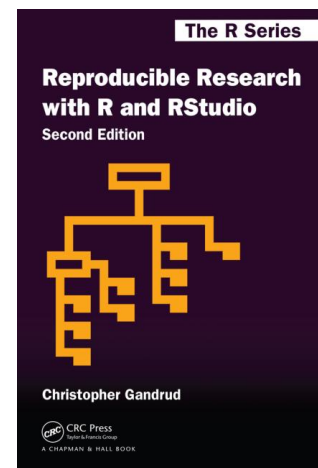
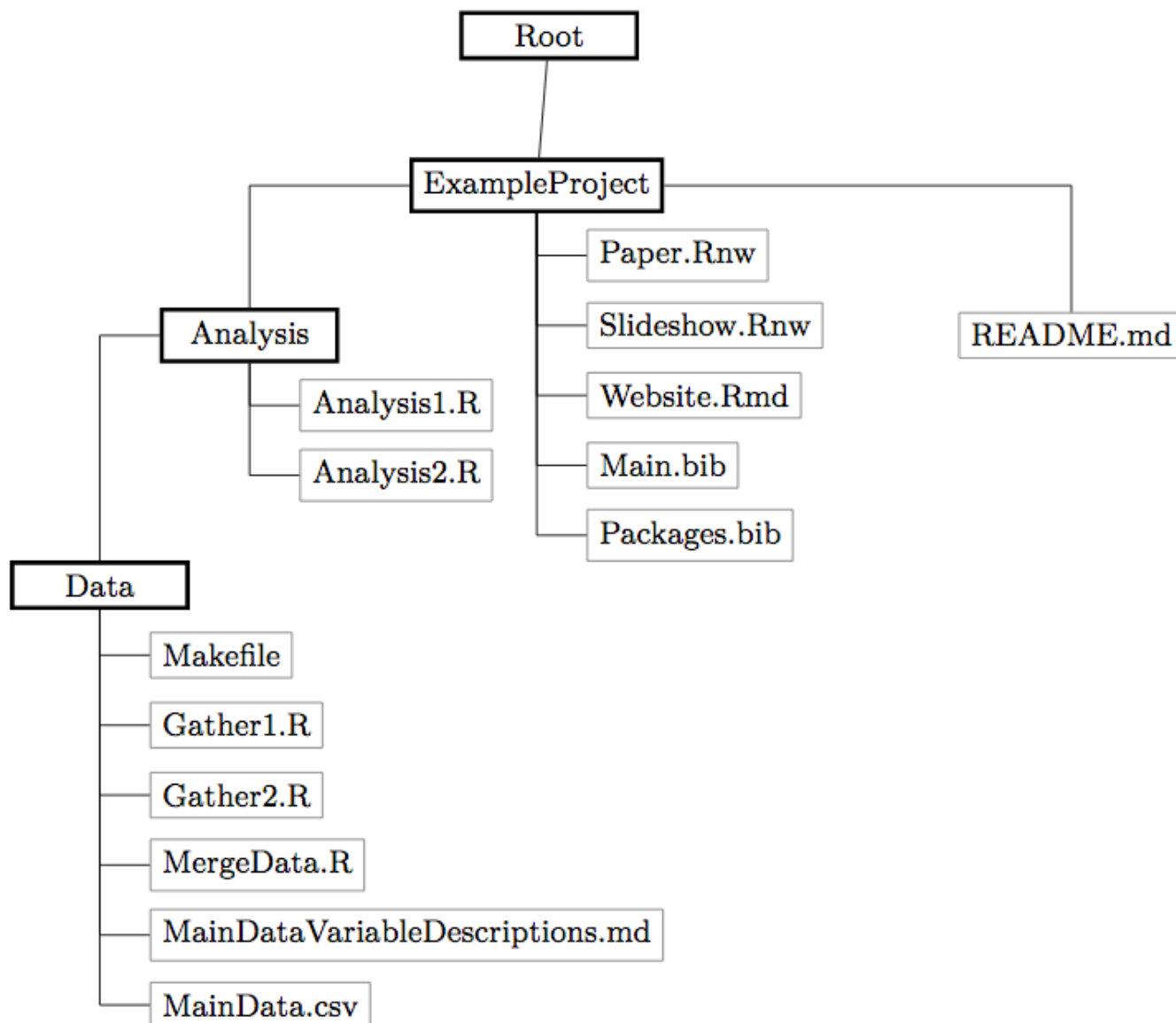
**/Code**

On the fly....

/Results

/Materials

FIGURE 4.1: Example Research Project File Tree



## **Comment your code**

# loading data

# variable transformation

# merging tables

# models for table 1

## Output

```
doedit "Z:\M14160\assessment\self data\data\stata operation-day 2-binary\EQ-I  
> Q-command-LR.do"
```

```
. do "C:\Users\lpxwh\AppData\Local\Temp\STD02000000.tmp"
```

```
. use EQIQbinaryLR.dta,clear
```

```
.  
end of do-file
```

```
. do "C:\Users\lpxwh\AppData\Local\Temp\STD02000000.tmp"
```

```
. logit BEQ2 gender representative age IQ
```

```
Iteration 0:   log likelihood = -157.28934
```

```
Iteration 1:   log likelihood = -155.59433
```

```
Iteration 2:   log likelihood = -155.59392
```

```
Iteration 3:   log likelihood = -155.59392
```

Logistic regression

Number of obs = 227

LR chi2(4) = 3.39

Prob > chi2 = 0.4947

Log likelihood = -155.59392

Pseudo R2 = 0.0108



```
#####
# load human rights data set and select only non-OECD countries;
# call the subset "t.nono"
#####
```

```
load("replicationdata.Rdata") # this data file needs to be in the folder of your working directory
t.nono <- replicationdata[replicationdata$oeecd==0, ] # select developing nations only (non-OECD)
t.nono <- droplevels(t.nono) # drop the unused levels (http://stackoverflow.com/questions/17217951/how-can-i-drop-unused-levels-from-a-data-frame)
```

```
#####
# Main models for paper
```

```
..
# Creating table "Total FDI and Personal Integrity Rights Protection (1983-2010)" in the main part of the paper ####
# US FDI ologit with robust standard errors clustered by country
CIRI.logit.LDV <- lrm(CIRI_PHYSINT ~ Lag_lognonnegUS_fdi_totalpGDP + Lag_logtrade + Lag_logGDPpc + Lag_logpopulation +
  Lag_polity2 + Lag_confl + Lag_CIRI_PHYSINT,data=data.CIRI.PHY.LDV,x=TRUE, y=TRUE)
CIRI.logit.LDV.corr <- robcov(CIRI.logit.LDV,data.CIRI.PHY.LDV$country) #robust errors clustered by country

PTS.logit.LDV <- lrm(PTS_ai_reversed ~ Lag_lognonnegUS_fdi_totalpGDP + Lag_logtrade + Lag_logGDPpc + Lag_logpopulation
  + Lag_polity2 + Lag_confl + Lag_PTS_ai_reversed,data=data.PTS.LDV,x=TRUE, y=TRUE)
PTS.logit.LDV.corr <- robcov(PTS.logit.LDV,data.PTS.LDV$country)
```

```
#####
#Creating "Correlation matrix (I) to (III)
#####
```

```
corstars <- function(x){
  x <- as.matrix(x)
  R <- rcorr(x, type="spearman")$r
  p <- rcorr(x)$P
  # n <- rcorr(x)$n
  mystars <- ifelse(p < .01, "****", ifelse(p < .05, "*** ", ifelse(p < .1, "* ", " ")))
  R <- format(round(R, 3))
  Rnew <- matrix(paste(R, mystars, sep=""), ncol=ncol(x)) #removed n here
```

```

148
149 First we load the data and R packages. The data file is part of the paper
    also made a copy available online.
150 <<message=FALSE,tidy=FALSE>>=
151 library(survival)
152 library(kernlab)
153 library(rms)
154 library(spatstat)
155 library(RColorBrewer)
156 library(gplots)
157
158 ## load data file from local copy or from URL
159 if (file.exists("Schwarz2015-supplement.Rdata")){
160   load("Schwarz2015-supplement.Rdata")
161   cat("Data loaded from local copy")
162 } else {
163   load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
164   cat("Data loaded from URL")
165 }
166
167 The first object in the .Rdata file is a table \texttt{D} with patient
168 <<message=FALSE>>=
169 D
170 attach(D)
171 @
172
173 <<echo=FALSE,message=FALSE>>=
174 ## Print the data table in LaTeX format for inclusion into main manuscript
175 library(xtable)
176 print(xtable(D),file="TableOverview.tex")
177 @
178
179 Rownames correspond to sample identifiers. Columns indicate the patient
    t (\texttt{Nr}), as well as values for temporal heterogeneity (\texttt{CE}),
    index (\texttt{CE}), overall survival in days (\texttt{OS}), progression
    (\texttt{PFS}) and indicators for survival (\texttt{dead}) and progression
    covariates for age, stage (ordered factor), residual disease after debul
    l), ordered factor) and the number of samples per patient (\texttt{N}).
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

## 1 Clinical data

### 1.1 Data overview

First we load the data and R packages. The data file is part of the paper supplement, and we have also made a copy available online.

```

library(survival)
library(kernlab)
library(rms)
library(spatstat)
library(RColorBrewer)
library(gplots)

## load data file from local copy or from URL
if (file.exists("Schwarz2015-supplement.Rdata")){
  load("Schwarz2015-supplement.Rdata")
  cat("Data loaded from local copy")
} else {
  load(url("http://www.markowetzlab.org/supplements/Schwarz2015-supplement.Rdata"))
  cat("Data loaded from URL")
}

## Data loaded from URL

```

The first object in the .Rdata file is a table D with patient information:

```

D
##      Nr      TH      CE      OS      PFS      dead      prog      Hist      Age      Stage      residual
## OV03-01  1  4.730231  1.2605274  511  271      1      1  HGSOC  47      IV      <1cm
## OV03-02  2      NA  0.7105901  977  363      1      1  HGSOC  62      IV      <1cm
## OV03-04  3  3.735366  1.2432629  209  153      1      1   HGS      69      IV      >1cm
## OV03-07  4      NA      NA      625  616      1      1  HGSOC  48     IIIC      Nil
## OV03-08  5  3.801712  1.4705531  547  303      1      1  HGSOC  63      IV      <1cm
## OV03-10  6  6.588895  0.7298828  744  298      1      1  HGSOC  59      IV      <NA>
## OV03-13  7  3.000290  0.6836961  1587  358      1      1  HGSOC  61      IV      >1cm
## OV03-17  8  3.423112  2.2357817  889  373      1      1  HGSOC  51     IIIC      <1cm
## OV03-20  9  4.487828  0.6494353  1278  563      1      1  HGSOC  71      IV      >1cm
## OV03-21 10  4.719848  0.8686309  1139  303      1      1  HGSOC  60     IIIC      >1cm
## OV03-22 11  5.702720  0.4834086  1556  382      1      1  HGSOC  58     IIIC      <1cm
## OV03-23 12      NA      NA     1565  534      1      1  HGSOC  60     IIIC      Nil
## OV03-24 NA      NA      NA      376  375      1      1  HGSOC  53     IIIC      >1cm
## OV03-25 13      NA  0.6215297  1166  776      1      1  HGSOC  57     IIIC      >1cm
## OV04-20 14  4.621984  0.6083119  1513  601      0      1  HGSOC  63     IIIC      Nil
## OV04-21 15      NA  0.7412773  706  332      1      1  HGSOC  54      IV      Nil
## OV04-27 16      NA      NA     1408  1408  0      0  HGSOC  58     IIIC      Nil
## OV04-30 17      NA  0.8591205  849  293      1      1  HGSOC  60     IIIC      >1cm
##      N

```

## **Methods section in paper**

Describe methods clearly

Name exact models with citations for statistical choices

Footnote should contain software versions

If space is restricted: Appendix

## Models

For the ordered categorical outcome variables, CIRI and PTS, I estimate an ordered logit model with robust standard errors clustered on country to correct for heteroskedasticity. For the continuous outcome variable, the Latent Human Rights Scores by Fariss, I employ ordinary least squares (OLS) with panel-corrected standard errors (PCSE).<sup>17</sup> In both models, I include a lagged dependent variable (“Past”) since countries that repressed their citizens in the past are more likely to use repressive acts in the future (Gurr 1988). A lagged dependent variable also corrects the serial correlation (Beck and Katz 1995, 2009).

I include a one-year lag between the outcome and predictors to allow the effect of FDI stock to spread in the country. This means that the accumulated FDI in a country in a given year is expected to correlate with rights protection in the following year, which establishes a time order and suggests a direction of causality from  $FDI(t-1)$  to rights protection( $t$ ).

The data set ranges from 1983 to 2010 and includes up to 121 non-OECD countries. The selection of these cases is limited to countries with available data on personal integrity rights and FDI measures (see a list of all countries in the online appendix). Two main sets of models are estimated: The first set of models includes total FDI to compare my results with previous work, while the second set of models replaces total FDI with investment in 10 industry sectors.

17. For the ordered logit models, I use the functions `lrm()` and `robcov()` from the R package “rms” Version 4.3-0, which produces the same results as the corresponding STATA command `ologit` with the `cluster()` modification (STATA Version 13.0). For OLS with PCSE, I use the functions `plm()` and `vcovBK()` from the R package “plm” Version 1.3.1, which produces the same results as the STATA command `xtpcse` with the pairwise specification.

## **Share your materials**

Readme file

Data, code, variable codebook

Information to reconstruct data from original sources

# Data sharing platforms



GitHub

---

UK Data Service

---



OSF



**HARVARD**

Dataverse

**ICPSR**

Inter-university  
Consortium for Political  
and Social Research

**CISER** **CORNELL INSTITUTE for**  
Social and Economic Research






 Metrics


50 Downloads

 Contact  Share

## Replication Data for: "Foreign Direct Investment and Repression: An Analysis Across Industry Sectors" Version 1.0

Janz, Nicole, 2017, "Replication Data for: "Foreign Direct Investment and Repression: An Analysis Across Industry Sectors"", <https://doi.org/10.7910/DVN/TT7NZJ>, Harvard Dataverse, V1, UNF:6:HfNaxyecbqKGLOf0hD3cDw==

 Cite Dataset ▾

 Learn about [Data Citation Standards](#).

### Description

The impact of foreign direct investment (FDI) on repression in developing nations is still disputed. Some argue that FDI improves economic development and exports human rights values. Others criticize the exploitation of cheap labor and resources, which may lead to tensions and government oppression. Previous studies have employed aggregate FDI data with conflicting results. Alternatively, I propose that the effects depend on what kind of FDI enters a country. I build a sectoral framework to discuss how skills and technology levels, as well as the motivation for FDI, can mediate the impact. I then examine the link in a panel data analysis (1983–2010) in 121 countries, integrating sectoral FDI in several resource, manufacturing, and service industries. The results show that investment in high-skilled and high-tech sectors has positive effects. The results are robust across several measures for repression, and when accounting for sector size, regional and time effects.

### Subject

Social Sciences

### Keyword

human rights, foreign direct investment, repression, sectors

Files

Metadata

Terms

Versions

#### 4 Files



[JHR\\_Replication.R](#)

R Syntax - 316.3 KB - Jun 5, 2017 - 7 Downloads  
MD5: b38d4241da685f4d3c42b7bbfde8a486



[JHR\\_Replication\\_Notes.pdf](#)

Adobe PDF - 70.0 KB - Jun 5, 2017 - 6 Downloads  
MD5: 10e97b3ebad5ea825e70559bb58e6c01



[replicationdata.tab](#)

Tabular Data - 1.8 MB - Jun 5, 2017 - 20 Downloads  
49 Variables, 5568 Observations - UNF:6:HfNaxyecbqKGLOf0hD3cDw==



[Variable\\_Codebook.xlsx](#)

MS Excel (XLSX) - 38.0 KB - Jun 5, 2017 - 17 Downloads  
MD5: 60199b9b4e380bd8b9c647eeaac9daf1

Code

Data

Variables



Replication Notes for  
Foreign Direct Investment and Repression:  
An Analysis Across Industry Sectors

**Nicole Janz** *University of Nottingham*

School of Politics and International Relations

University Park

Nottingham, NG7 2RD

Email: [nicole.janz@nottingham.ac.uk](mailto:nicole.janz@nottingham.ac.uk)

September 14, 2016

# Instructions for Replication

- First, load the data. This is a panel data set with country-years. Each table or figure is identified by its header as in the article or online appendix. If you are looking for a specific table, search the (admittedly long) Rscript for that particular heading.
- Run the code chunk by chunk. Many variables are created 'on the go' and re-used at a later point. Make sure to run the code in the same order, and run all the code even if you only want to replicate e.g. the last table.
- In order to preserve the largest possible sample size, I have created data sets for each model separately (based on the original table that you have loaded in R). This way, depending on the particular model, slightly different country-years are included. Make sure to create all these data sets (again, in order of the code) to run the models.

## Citation

If you work with the data for your own study (replication or original work based on these data), please cite my article as well as the data set. It would be great if you could let me know about your results. A suitable citation of the data is provided by Dataverse where you downloaded the data.

## Contact

Please contact me if you have any questions about the study or replication files at: [nicole.janz@nottingham.ac.uk](mailto:nicole.janz@nottingham.ac.uk) or [nicolejanz@gmail.com](mailto:nicolejanz@gmail.com). Any feedback on your replication attempts is more than welcome.

## Acknowledgements

Gu Li and Sergio Cuesta have verified that the provided Rscript produces the tables and figures for the study. All errors remain my own.

**Data  
transparency**

**Providing full  
access to data  
itself**

**Analytic  
transparency**

**Information  
about data  
analysis**

**Production  
transparency**

**Process of data  
collection**

## **Quantitative research**

Upload datasets  
used for analysis

Code for models  
(SPSS, STATA, R)

Good methods  
section in paper

Provide or describe  
raw data & variable  
codings

**Replication  
Exercises**



**Transparent  
workflows**



# Five Selfish Reasons to Share Data

## Quality

Establish trust & credibility in your work

## Reputation

Be known as a transparent researcher in your field

## Citation

Your data will be cited

## Consistency

Transparent workflow makes it easier to re-use your own data later

## Practicality

Meet journal & funder standards



Photo: Sam Mather

# Thank you!



[nicole.janz@nottingham.ac.uk](mailto:nicole.janz@nottingham.ac.uk)



[@polscireplicate](https://twitter.com/polscireplicate)



Political Science Replication Blog





## Data transparency

Providing full access to data itself

## Analytic transparency

Information about data analysis

## Production transparency

Process of data collection

# Qualitative research

Provide (partial)

- Interview transcripts
- Field notes
- Videos...

Describe which evidence supports which claims

Discursive footnotes / supplement

Explain how data were collected:

- Interviewee selection
- Participants
- Documents...

# Literature on Replication

- King, Gary. (2006). **How to Write a Publishable Paper as a Class Project**, copy at: <http://gking.harvard.edu/papers>
- Janz, N. (2015) **Bringing the Gold Standard Into the Class Room: Replication in University Teaching**, International Studies Perspectives, Article first published online: 9 March 2015. Copy at: <http://tinyurl.com/q2qnrvn>
- Brandt et al. (2014) **The Replication Recipe: What makes for a convincing replication?** Journal of Experimental Social Psychology, Vol 50, pp. 217-224. Copy at: <http://tinyurl.com/poe474k>
- Markowetz, Florian (2015), **Five selfish reasons to work reproducibly**. Genome Biology 16:274.

# Literature on Transparent Workflow

- Christensen, Garret (2016). **Manual of Best Practices in Transparent Social Science Research**  
<https://github.com/garretchristensen/BestPracticesManual>
- Open Science Framework. **Transparency and Openness Promotion (TOP) Guidelines**. <https://cos.io/top/>
- **TIER Documentation Protocol**  
<https://www.haverford.edu/project-tier/protocol-v2>
- Janz, Nicole & Figueiredo, Dalson (2017, March 13). **Workshop: The Gold Standard of Reproducible Research**. Retrieved from <https://osf.io/2fqnw/> (slides, handouts)

# Literature on Political Science Debate

Moravcsik, A. (2014). **Transparency: The Revolution in Qualitative Research.** *PS: Political Science & Politics*, 47(1), 48-53.  
doi:10.1017/S1049096513001789

Lupia, A., & Elman, C. (2014). **Openness in political science: Data access and research transparency.** *PS - Political Science and Politics*, 47(1), 19-42. doi: 10.1017/S1049096513001716